DOCUMENT RESUME

ED 034 235 CG 004 606

AUTHOR	Campbell, Joel T.; And Others
TITLE	Prediction of Job Performance for Negro and White Medical Technicians. A Regression Analysis of Potential Test Bias: Predicting Job Knowledge Scores from an Aptitude Battery.
INSTITUTION	Civil Service Commission, Washington, D.C.; Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Ford Foundation, New York, N.Y.
PUB DATE	Apr 69
NOTE	16p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.90
DESCRIPTORS	*Aptitude Tests, Caucasians, Cultural Factors, Employment Qualifications, Ethnic Groups, Job Skills, Minority Groups, Negroes, *Performance Criteria, *Prediction, Predictive Validity, Testing Problems, *Test Validity

ABSTRACT

An aptitude battery was validated against a Job Knowledge Test to determine if aptitude tests are culturally biased when used to predict occupational success. 287 white and 168 Negro Medical Technicians completed an eight test experimental aptitude battery and a Job Knowledge Test developed as a criterion measure of job performance. Linear regression analysis was used to detect the potential bias. All the aptitude tests were found to be culturally biased, but seven of them displayed bias in the direction opposite to that expected. For any predictor score on the seven tests, white technicians were likely to score better on the Job Knowledge Test than did Negroes with the same score. Consequently the bias was against the whites in that they performed better on the criterion for any given predictor score. To the investigators' knowledge there is no hypothesis or theory to explain the phenomenon of reverse unfairness which they have encountered. (RM)

CG 004 606

PR-69-6

PREDICTION OF JOB PERFORMANCE FOR NEGRO AND WHITE MEDICAL TECHNICIANS

A Regression Analysis of Potential Test Bias: Predicting Job Knowledge Scores from an Aptitude Battery

Joel T. Campbell, Lewis W. Pike,
and Ronald L. Flaugher

April 1969

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

ED034235

Personnel

Management Committee

Albert P. Maslow, Chief, Personnel Measurement Research and Development Center, U.S. Civil Service Commission

William W. Turnbull, Executive Vice-President, Educational Testing Service

Advisory Committee

John K. Hemphill, Far West Laboratory for Educational Research and Development, Chairman

Marvin D. Dunnette, University of Minnesota

Robert M. Guion, Bowling Green State University

S. O. Roberts, Fisk University

Project Staff

From U.S. Civil Service Commission

William A. Gorham, Program Director

From Educational Testing Service

Joel T. Campbell, Senior Research Psychologist, Principal Investigator

Ronald L. Flaugher, Research Psychologist

Lewis W. Pike, Research Psychologist

Donald A. Rock, Research Psychologist

Franklin R. Evans, Associate Research Psychologist

Mrs. Lila Norris, Associate Research Psychologist

Mrs. Virginia Rau, Administrative Assistant

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

PROBLEM[1]

A meaningful answer to the question of whether selected aptitude tests will be biased when used to predict success in a particular occupation depends in large measure on the selection of an appropriate criterion. Supervisory ratings, a test of job knowledge, and work-sample tests are among the more likely criteria to be used. Relationships between the first two are examined in this paper, to provide information about the comparative strengths and weaknesses of each for detecting such selection bias. Data for this examination came from a study of Medical Technicians working in Veteran's Administration Hospitals. These data were particularly useful for a validation study, because the great majority of technicians had <u>not</u> been selected on the basis of aptitude or achievement tests, thus avoiding the perennial problem of restriction of range.

In a previous paper (Flaugher, et al., 1969) the authors described the difficulties encountered when attempting to employ a supervisory rating scale as a criterion against which to validate paper and pencil aptitude tests for the job of medical laboratory technician. To summarize those findings briefly, it was discovered that the relationship between the predictor tests and the criterion of supervisory ratings was moderated dramatically by the interaction of the ethnic identities of the person completing the rating and the person receiving the rating. When attempting to raise the question of whether or not the potential selection tests were biased against any ethnic group, therefore, it became clear that the question was meaningless for an overall relationship between the tests and ratings. Rather, four predictor-criterion relationships had to be considered, one for each combination of rater and ratee's ethnic identity.

---

[1] This study was funded by the Ford Foundation and conducted jointly by the Educational Testing Service and the United States Civil Service Commission.

However, if there are four kinds of relationship between a predictor test and supervisory ratings, then four separate validity coefficients must be determined. Further, it is very unlikely that the typical employment situation will have enough applicants or incumbents to safely subdivide their validation samples to that extent, so the whole practice of using supervisory ratings as validating criteria is thrown open to serious question. On the other hand, some kind of criterion measure must be used if test validity is to be measured and the problem of test bias is ever to be understood.

For some occupations, a viable alternative to using supervisory ratings as criteria of job performance may be to employ an objective paper and pencil test of the information important for performing the job adequately. To the extent that an occupation requires technical information, particularly information that may be learned on the job as well as through formal instruction, such an objective measure may be a very suitable criterion. Where it is appropriate, this alternative has particular appeal in that it circumvents the effects of biased ratings described in the paragraphs above. An objection can be raised to the use of a job knowledge test as the performance criterion, however, because of what has been described as "criterion contamination" by Brogden and Taylor (1950). As it applies to the present problem, this concept refers to the possible existence of a factor not related to job success which influences both predictor and criterion measures, thus spuriously inflating the obtained validity coefficient. In the present instance, the concern would be that Negro incumbents will tend to be less experienced in test-taking than their white peers, with the result that their scores on both predictor and criterion will be depressed by some amount, reflecting neither their true aptitude nor their actual knowledge of the job.

Another alternative criterion measure, that of a job sample, or work tryout, would appear to meet the criticisms of either ratings or job knowledge tests, but the fairness and adequacy of this type of criterion measures is more apparent than real.

The nature of the occupation of Medical Technician is such that a job knowledge test is especially likely to be a suitable performance criterion. The occupation is characterized by the frequent use of an unusually high degree of technical information and of a sizable amount of written material. Thus a written examination covering the relevant information for this occupation would appear appropriate as at least a partial index of competence on the job.

Because of these considerations, a Job Knowledge Test was developed as a criterion measure, against which a number of aptitude tests were validated in a search for evidence of test bias. Linear regression analysis was used to detect potential bias, using a number of experimental aptitude tests in turn as predictors of the Job Knowledge Test criterion.

PROCEDURE

The development of the 63-item Job Knowledge Test is described in detail elsewhere (Pike, 1969). A total of 287 white and 168 Negro Medical Technicians completed both an eight-test aptitude battery and the Job Knowledge Test. Means, variances, and inter-correlations were calculated, and regression lines were determined for each aptitude test against the Job Knowledge Test.

RESULTS

Means and standard deviations are presented in Table 1 separately by race for the Job Knowledge Test "criterion" and for each of the predictor measures.

Also included are the individual correlations of each of the predictors with the Job Knowledge Test and the t-test significances between the means. It is seen that Negroes scored significantly lower as a group on all measures, that the variances are smaller for the Negroes on all predictors, but not on the Job Knowledge Test, and that the correlation of each predictor measure with the Job Knowledge Test is larger for the Negro group.

Using the data from Table 1, regression lines for the two groups of incumbents were graphed for each predictor against the Job Knowledge Test. These appear in Figures 1 to 8. The results of the Gulliksen-Wilks analysis of covariance test for significance between regression lines (Gulliksen and Wilks, 1950) appears in the upper right hand corner of each figure. In all eight cases there was no significant difference between the errors of estimate around the regression lines, indicating that both whites and Negroes have approximately equal dispersion of test scores. In seven of the eight cases, there was no significant difference between the slopes of the regression lines, indicating that these seven tests predicted Job Knowledge Test scores equally well (or equally poorly) for the two ethnic groups. The slopes of the regression lines are significantly different for the Necessary Arithmetic Test (Figure 4), indicating that this test predicts Job Knowledge Test scores better for Negroes than for whites.

Since the regression line slopes for predicting the criterion from Necessary Arithmetic scores are not equal for the two ethnic groups, it is not appropriate to test the difference in intercepts. For the other seven aptitude tests, however, there was a significant difference in intercepts, with the white group having the higher intercept in each case.

Discussion

Figures 1-8 show graphically some of the problems encountered when

attempting to determine whether or not a given test contains cultural or ethnic bias. A possible interpretation of the regression lines in Figure 4, for example, is the following: White applicants who receive a score below 13 on the Necessary Arithmetic test, typically perform better on a criterion of job knowledge than Negroes receiving the same score. Since the score on the predictor test is the same, and the criterion scores are different, the predictor test may be said to be biased. In this case, the bias is against the white applicant, in that he performs better on the criterion for any given predictor score. For this same test, however, white applicants receiving a score above 13 typically do not perform as well on the criterion as do Negroes receiving the same score. In the latter instances, the test as a selection device may be said to be biased against the Negro applicant.

Thus, this particular test may be said to contain bias, but the direction (and magnitude) of the bias differs, depending upon the score range being considered.

In an employment situation which employed a cutoff score for hiring or promotion, the minority group applicant might be at a disadvantage, depending upon the cutoff score used. For example, if the cutoff score on the Necessary Arithmetic test (see Figure 4) was set at 20, then whites would be hired who would eventually be expected to reach about 38 on the criterion measure. But, if hiring is on the basis of an expected minimal performance of about 38, then Negroes scoring between 16 and 20 on the Necessary Arithmetic Test have been unfairly passed over, since they, too, would be expected to reach that criterion level. In effect, then, more test performance would have been demanded of these Negroes than is necessary for them to reach the desired minimum criterion.

In a merit system, where selection is from the top, Negroes would be at an even greater disadvantage, if only a relatively small proportion of candidates for promotion were to be advanced. In working downward through a list of Necessary Arithmetic scorers, a substantial number of whites would probably be selected before the highest scoring Negroes would be reached. Yet many of the whites selected would be predicted not to do as well on the criterion measure as would the best Negro scorers on the predictor.

The regression lines also cross for two other tests, Extended Range Vocabulary and Subtraction and Multiplication. However, since the co-variance analysis indicates that the slopes of the regression lines are not significantly different, it is not appropriate to make similar detailed interpretation. For both of these tests, however, and for the remaining five tests in the battery, the intercepts of the regression lines are different. It is therefore appropriate to interpret the difference in <u>level</u> of all seven pairs of regression lines. In all seven cases, then, the tests may be said to display bias, but in the direction opposite to that which would be expected. For any given predictor score, white technicians are likely to score better on the criterion than do Negro technicians, rather than vice versa.

An example may serve to clarify this interpretation. Say that a cutoff score of 25 on the Picture Number test (Figure 7) was one basis for hiring. Using this cutoff, some Negroes would be selected whose eventual expected score on the criterion would be about 33. But if a criterion score of 33 is acceptable, then white candidates scoring between 5 and 25 on the Picture Number test would have been unfairly excluded from consideration, since their expected criterion performance equals or exceeds the minimum level. And that group constitutes about half the entire white candidate population.

How is one to interpret these results? Certainly the usual hypothesis of cultural bias is not tenable, in that those persons with supposedly the more favorable background are at a disadvantage on a majority of the tests, were they to be used as selection devices.

As was pointed out earlier in this paper, one criticism that could be brought against this study is the particular criterion employed, that of "still another test," serving only to predict the biases existent in the criterion by using those biases existent in the predictors. This argument is somewhat weakened by the nature of the two types of tests employed here. In theory, test bias exists because the test contains cultural factors which are not relevant to the job performance to be predicted and which the minority groups have not had the opportunity to acquire. The Job Knowledge Test, however, is certainly measuring content more nearly related to the occupation than are the aptitude tests, and the minority group sample has had, on the average, several years experience on the job. Thus, the Job Knowledge Test, if biased at all, should be <u>less</u> biased than the aptitude tests, and its use as a criterion should help to expose that bias exists in the aptitude tests. The results of the present study, however, not only fail to expose the type of bias which would ordinarily be predicted, but in fact, presents some evidence for the existence of what might be called a reverse unfairness. To the authors' knowledge, no hypothesis or theory exists to explain this phenomenon.

References

Brogden, H.E., & Taylor, E.K. The theory and classification of criterion bias. Educational and Psychological Measurement, 1950, 10, 159-186.

Flaugher, R.L., Campbell, J.T., & Pike, L.W. Prediction of job performance for Negro and white Medical Technicians: Ethnic group membership as a moderator of supervisor's ratings. Project Report 69-5. Princeton, New Jersey: Educational Testing Service, 1969.

French, J.W., Ekstrom, R.B., and Price, L.A. Kit of reference tests for cognitive factors. Princeton, New Jersey: Educational Testing Service, 1963.

Gulliksen, H., & Wilks, S.S. Regression tests for several examples. Psychometrika, 1950, 15, 91-114.

Pike, L.W., Prediction of job performance for Negro and white medical technicians: Development of the instrumentation. Project Report 69-4. Princeton, New Jersey: Educational Testing Service, 1969.

Table 1

MEANS AND STANDARD DEVIATIONS OF RIGHTS SCORES BY ETHNIC GROUP

Measure	Number of items	N=approx. 168 Negro Mean	S.D.	r with JKT	N=approx. 297 White Mean	S.D.	r with JKT	t
"Criterion"								
Job Knowledge Test	63	31.4	10.4		35.7	9.4		4.40*
Predictors								
1. Subtraction and Multiplication	120	46.6	17.1	.34	57.0	19.6	.23	5.73*
2. Vocabulary	48	19.2	7.7	.32	25.7	9.1	.27	7.86*
3. Hidden Figures	32	5.7	4.0	.15	7.8	4.9	.14	4.48*
4. Necessary Arithmetic	30	10.4	3.7	.46	14.0	4.8	.34	8.63*
5. Number Comparison	96	36.5	9.7	.23	42.4	9.8	.14	6.24*
6. Gestalt Completion	20	10.3	5.2	.25	12.1	5.6	.17	3.38*
7. Picture-Number	42	17.5	8.8	.21	20.8	9.4	.16	3.77*
8. Paper Folding	20	6.8	3.3	.22	9.1	3.8	.21	6.55*

* significant at .01 level

Errata Sheet for PR-69-6

Pages 10 through 13

The following Gulliksen-Wilks chi square data were omitted from Figures 1 through 8:

Figure	Error of Estimate	Slope	Intercept
1	1.01	3.45	9.40 **
2	1.58	1.96	5.26 *
3	2.19	0.28	14.93 **
4	0.37	8.49 **	1.92
5	1.55	1.37	11.67 **
6	1.58	1.35	15.49 **
7	1.90	0.64	15.34 **
8	2.17	0.36	9.76 **

* $p < .05$

** $p < .01$

For all figures, the following legend applies:

———— Negro

- - - - - - White

△ Mean

• Standard deviation intervals

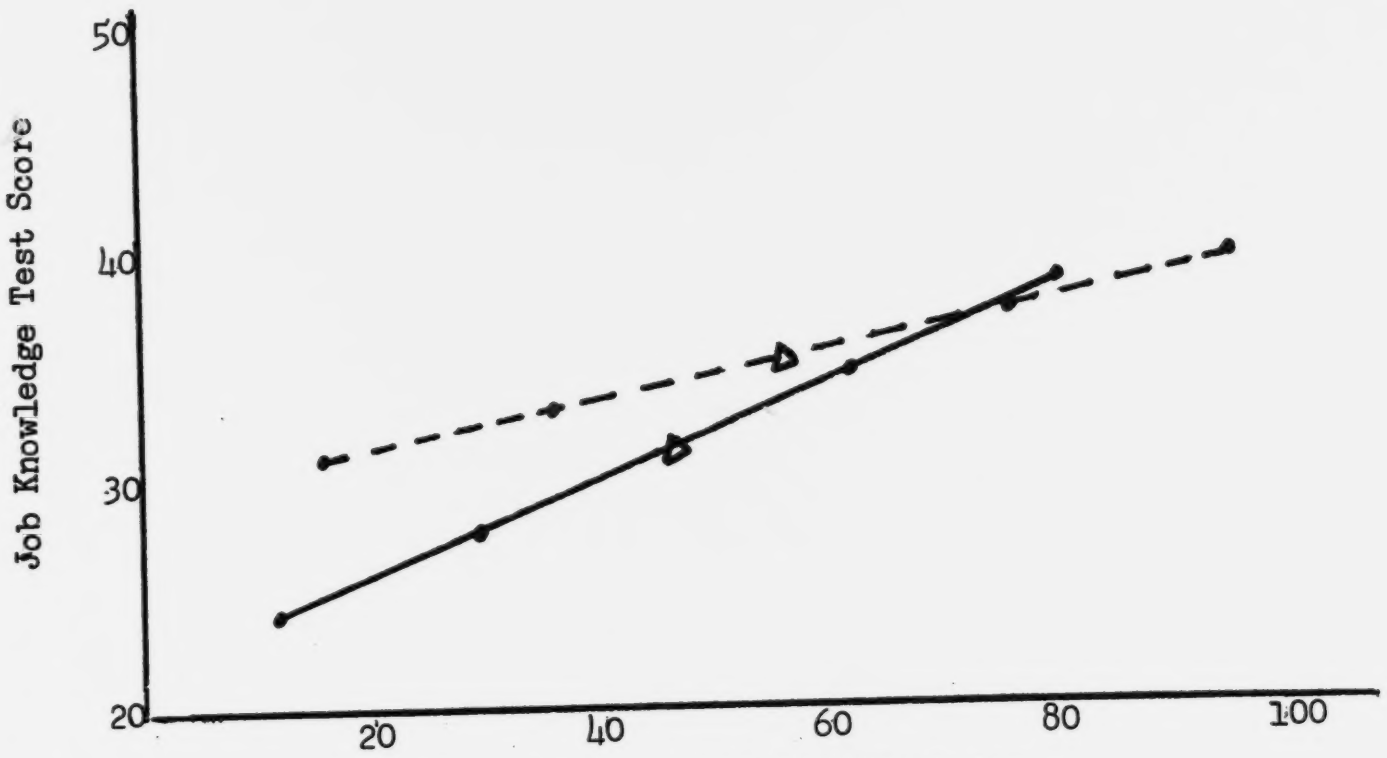

Fig. 1. Regression lines for the prediction of Job Knowledge from Subtraction and Multiplication Test scores, separately for Negroes and whites.

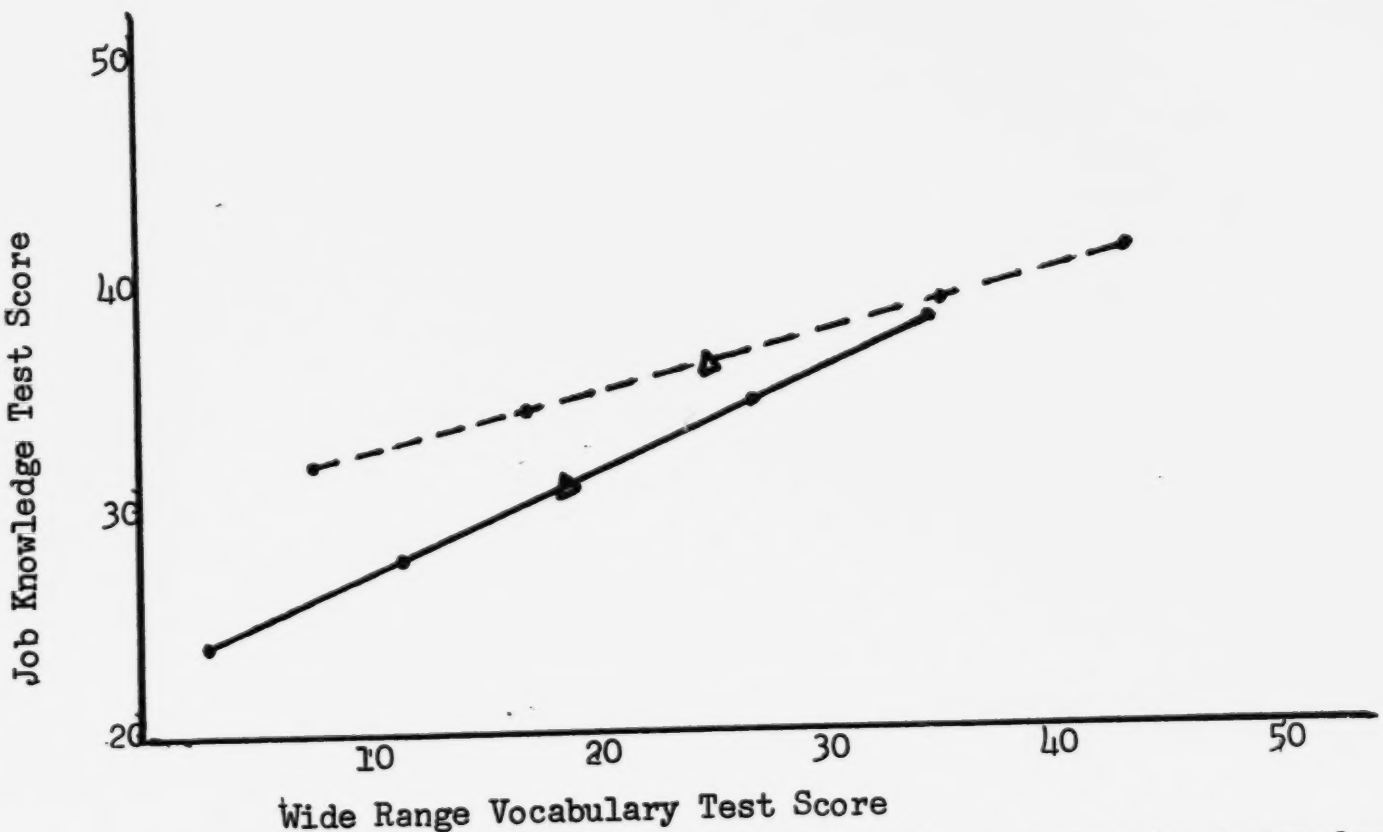

Fig. 2. Regression lines for the prediction of Job Knowledge from Wide Range Vocabulary Test scores, separately for Negroes and whites.

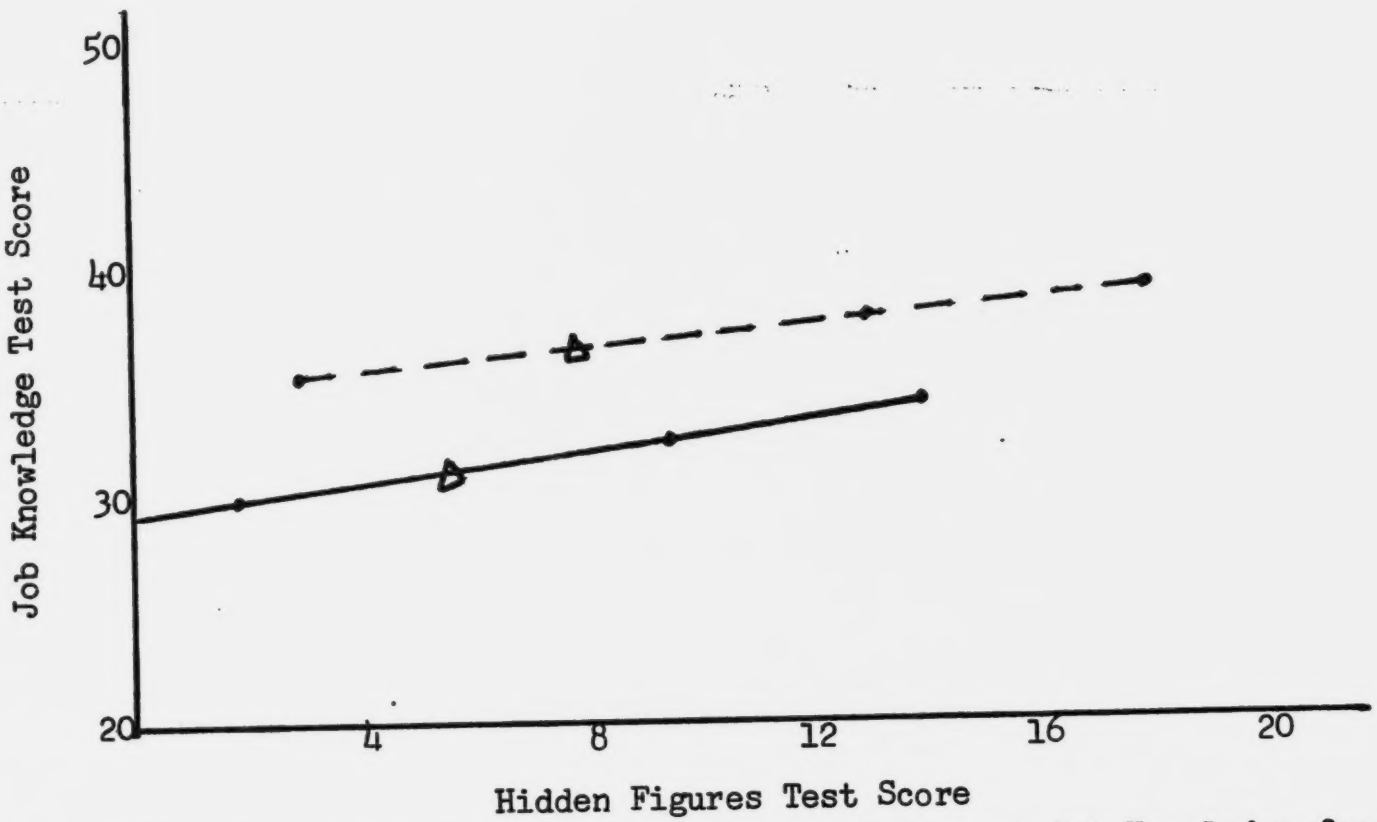

Fig. 3. Regression lines for the prediction of Job Knowledge from Hidden figures Test scores, separately for Negroes and whites.

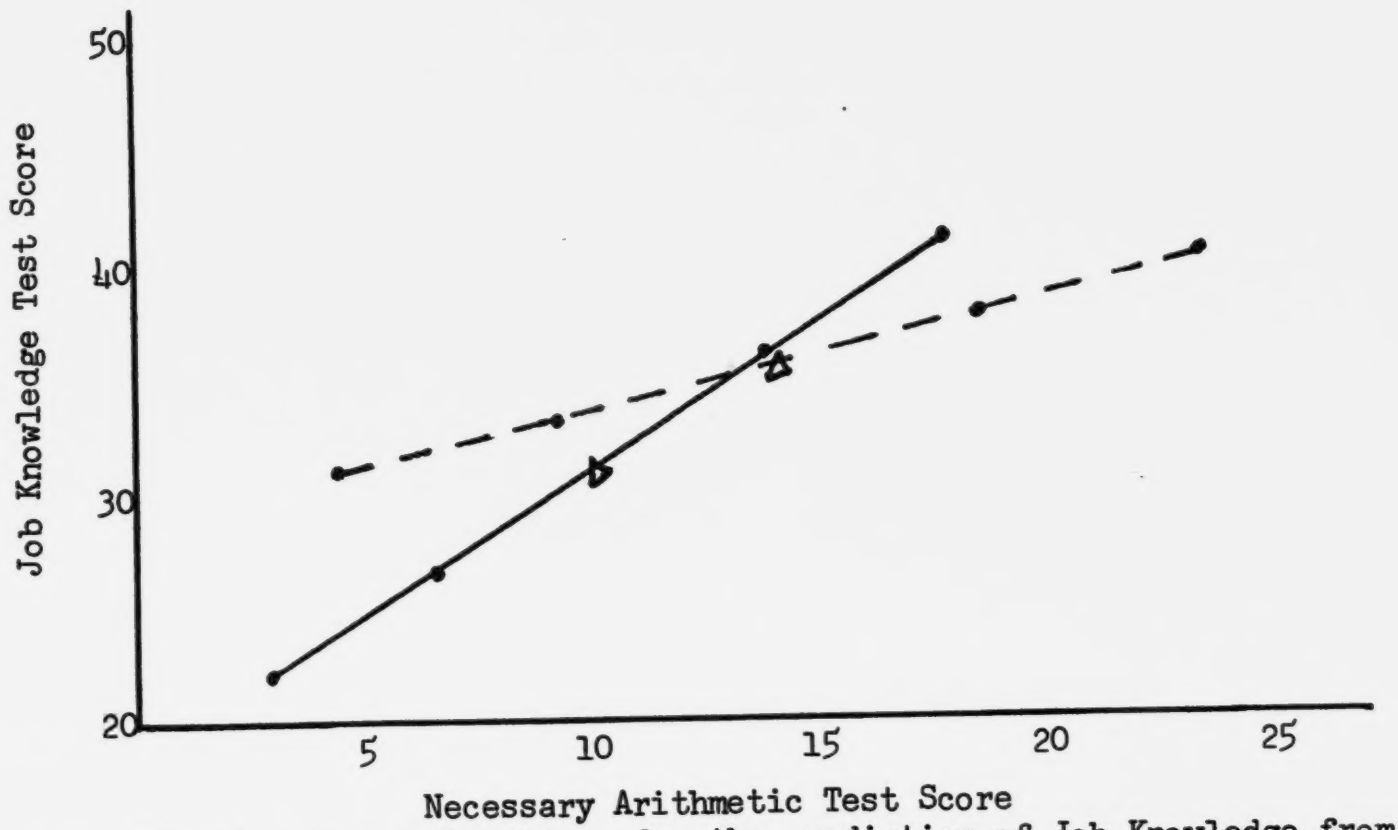

Fig. 4. Regression lines for the prediction of Job Knowledge from Necessary Arithmetic Test scores, separately for Negroes and whites

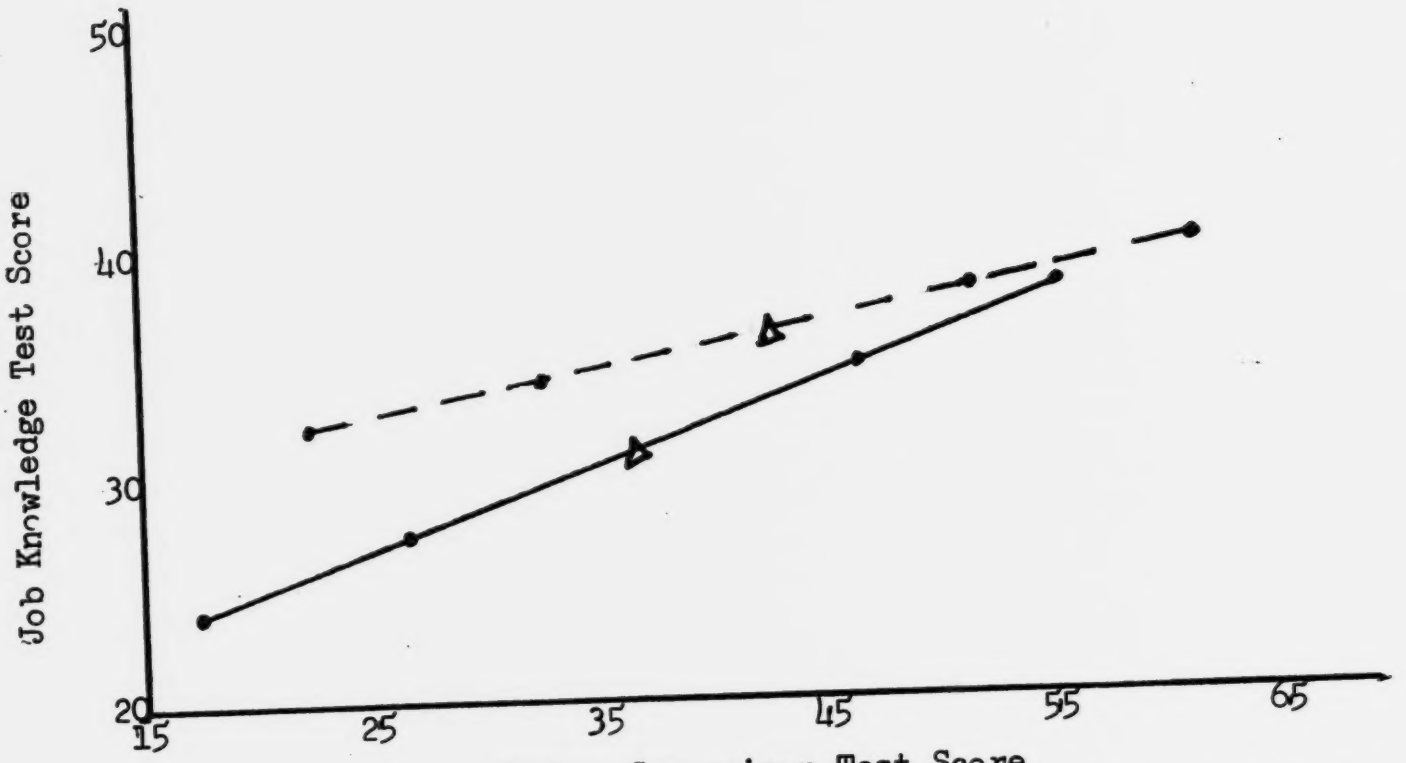

Fig. 5. Regression lines for the prediction of Job Knowledge from Number Comparison Test scores, separately for Negroes and whites.

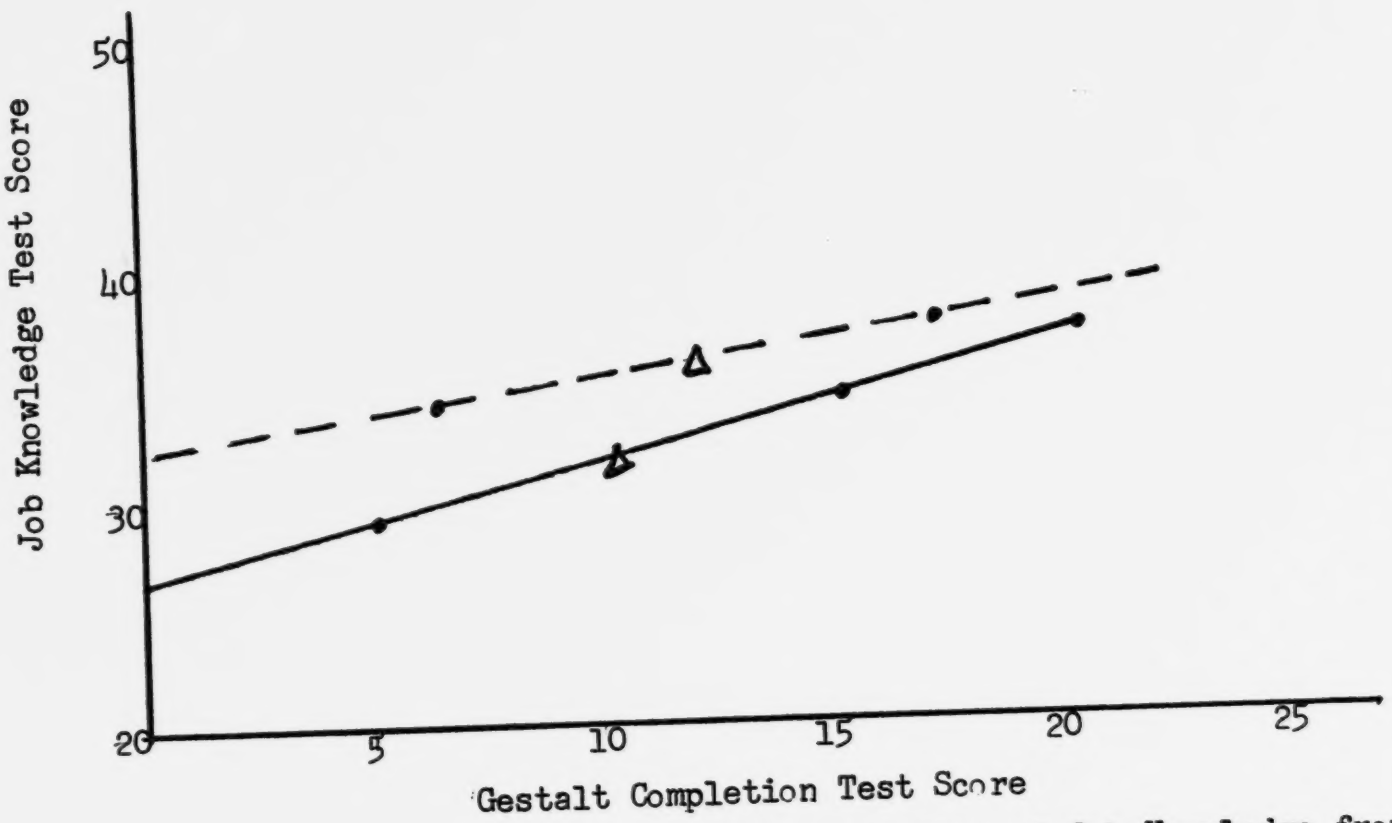

Fig. 6. Regression lines for the prediction of Job Knowledge from Gestalt Completion Test scores, separately for Negroes and whites.

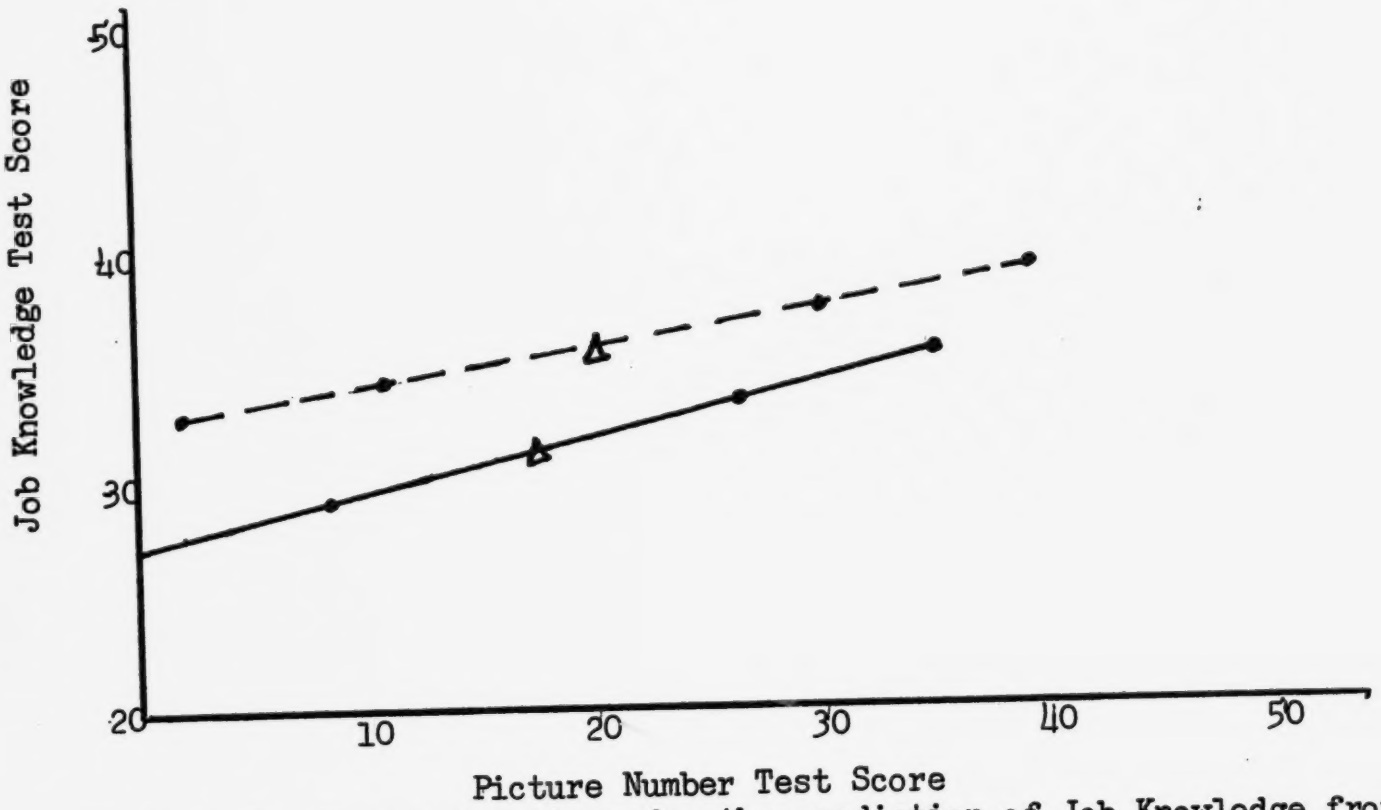

Fig. 7. Regression lines for the prediction of Job Knowledge from Picture Number Test scores, separately for Negroes and whites.

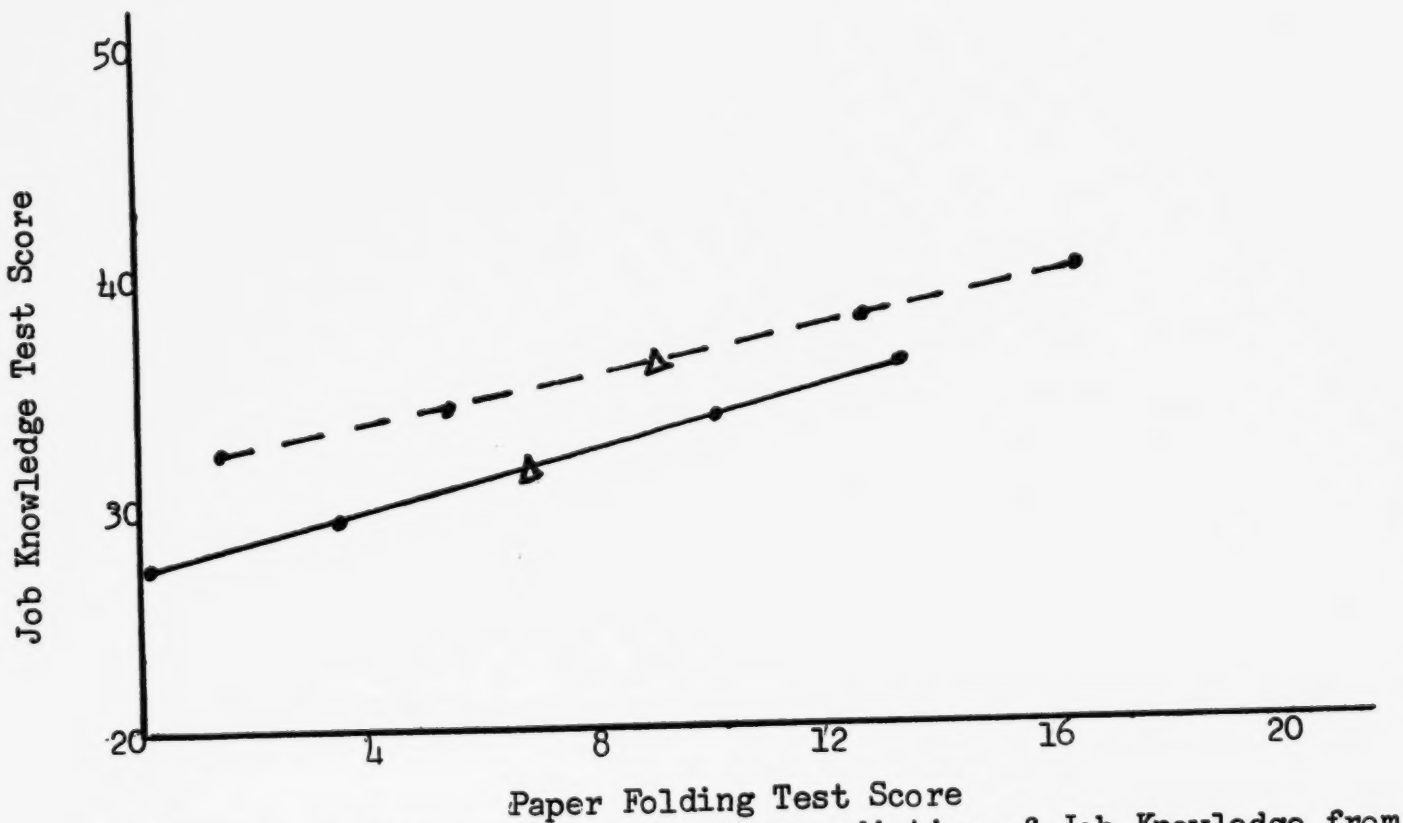

Fig. 8. Regression lines for the prediction of Job Knowledge from Paper Folding Test scores, separately for Negroes and whites.